Self-Learning Social Media Systems Through Machine Learning and Data Science

Christian Puzicha

KMWorld, Washington DC, 2018
About Raytion

15+ years of experience in full-service search & big data solutions

Large-scale search solutions for Global 500 since 2001

Professional services teams for entire project life cycle

Complementary search products for efficient solution delivery including connectors

Advanced Applications

Content Integration
Social Media Monitoring Solutions
Kendall Jenner Pepsi Ad – Social media crisis

Successful Ad
Supermodel’s first time collaboration with Pepsi
Famous public figure – Kendall Jenner
Social media campaigns promoting the Ad
Teasers on social media for the Ad
Kendall Jenner said: „Life opportunity“

Tone-deaf idea
Unhappy with the portrayal on civic disobedience on display.
Incorrectly re-enacting the Black Lives Matters protests
 Appropriating important protest movements in order to sell cola.

Social Media Spikes: 4 April – 6 April 2017
@KendallJenner is the first model since @CindyCrawford to have a @pepsi Global campaign

#PepsiKendall

So proud of you kenny!! ❤️ @KendallJenner

#PepsiKendall

@KendallJenner Apology to Kendall = ridiculous. Like screenwriter apologizing to actor for being in bad movie after reading script.

@KendallJenner

You did it kenny!

You did it kenny!

@pepsi

Lol @pepsi shut up. Don’t talk to her. Donate the fee you paid for that failure of a commercial to a cause that you can’t write off.
Sentiment towards Pepsi plummeted after ad release

Social data can be monitored BEFORE, DURING and AFTER crisis

Mentions of Pepsi soared after new ad

Most common phrases around Pepsi after ad release

Source: Brandwatch
How can you monitor social media and prevent the social media crisis?
Project use case – Telecommunication Company Social Media Monitoring Solution
Project use cases – Telecommunication company social media monitoring solution

Solution should provide live details about what is happening on social media for this brand and make these details easily actionable.

**Industry**
Telecommunication company

**Users**
Italian
(possibility to extend to other languages)
Key capabilities

Input data
- Facebook posts
- Facebook comments
- Users

Analysis
- Sentiment Analysis
- Italian Natural Language Processing
- Text analysis
- Key influencers
- Flexible Dashboards

Actions
- Alerts to social media team
- Response to customers
Challenges
General Challenges for Training a Sentiment Model

1. Non-English language

2. Short communication (max 280 characters)

3. Unbalanced input

4. Lots of typos, abbreviations, slang

5. Irony, sarcasm
Solution Description
Our solution
Our solution
Analyzed: #enterprisearch

@raytion_com
5 days ago
@raytion_com loves #search and #BBQ. Also this year, Raytion proudly sponsors the @berlinbuzzwords! You can meet us at our booth at the #b360 Get Together Monday, June 11, 6pm. Looking forward to talking to you! #berlin2018 #raytion #BBQ #enterprisearch

more details
Why search engine for solution

Out of the Box Features:

Connectors (or at least development kits)
Language Detection
NLP
Entity Extraction
Sentiment Analysis
UI including data manipulation
Data Storage
Search
Connect multiple sources into one system: homogenous system, consistent text analytics
Performant index
Flexible dashboards
Machine Learning Details
Using Sentiment Analysis – How it Works

1. **LABEL**
   Manual tagging for positive / negative

2. **TRAIN**
   Train the model with labeled data

3. **CLASSIFY**
   Ingest all data, including real-time new data and classify against the model

4. **MONITOR/RETRAIN/UPDATE**
   Optimize quality and tune model by labeling, retraining and classifying data on the run
Machine learning details

Algorithm

Feature generation „Bag of words“ on bi-grams

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>I will eat your donuts</td>
<td>^I, I will, will eat, eat your, your donuts, donuts$</td>
</tr>
<tr>
<td>I will see you eat donuts</td>
<td>^I, I will, will see, see you, you eat, eat donuts, donuts$</td>
</tr>
</tbody>
</table>

Optimization function: Perceptron (Linear predictor function)

\[ f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases} \]

Machine learning details

Validation, number of test cases

Cross-validation

<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt; fold</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; fold</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt; fold</th>
<th>4&lt;sup&gt;th&lt;/sup&gt; fold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Train</td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>Test</td>
<td>Test</td>
<td>Train</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>Test</td>
<td>Test</td>
<td>Train</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>Test</td>
<td>Test</td>
<td>Train</td>
</tr>
</tbody>
</table>

Success measures

Confusion Matrix

<table>
<thead>
<tr>
<th></th>
<th>Pos</th>
<th>Neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected as Pos</td>
<td>Correct</td>
<td>False positive</td>
</tr>
<tr>
<td>Detected as Neg</td>
<td>False negative</td>
<td>Correct</td>
</tr>
</tbody>
</table>
Machine learning details

Facebook data set
Customer specific
850 Facebook posts/comments
160 annotated samples
Italian

<table>
<thead>
<tr>
<th></th>
<th>Pos</th>
<th>Neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected as Pos</td>
<td>56%</td>
<td>23.08%</td>
</tr>
<tr>
<td>Detected as Neg</td>
<td>44%</td>
<td>76.92%</td>
</tr>
</tbody>
</table>

Twitter data set
Benchmarking dataset
1.6 million tweets
All annotated
English

<table>
<thead>
<tr>
<th></th>
<th>Pos</th>
<th>Neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected as Pos</td>
<td>81.51%</td>
<td>20.90%</td>
</tr>
<tr>
<td>Detected as Neg</td>
<td>18.49%</td>
<td>79.10%</td>
</tr>
</tbody>
</table>
Conclusion

1. Different content sources -> different models
2. Different language -> different models
3. Features of search engines are useful for social media monitoring
4. Success depends on size and quality of training data
5. Ethics, what to monitor and should I do it?
Questions and Comments?

Christian Puzicha
christian.puzicha@raytion.com